# Data Analysis in Social Science 3
# (SOC2094/3094 / POL2094/3094)

### Dr Alexey Bessudnov

### January 2018

## 1    Practical arrangements

**Classes:**

- Monday, 11.30am - 1.30pm, WSL 220

Classes begin on 15 January and end on 26 March. The class on 22 January (week 2) has been cancelled and moved to Friday 2 February, 10.30 - 12.30, WSL 220.

**Office hours (Amory A341):**

- Monday, 2-3pm

- Friday, 12-1pm

**Email:**

- A.Bessudnov@exeter.ac.uk

## 2    Aims of the module

This is a fourth module in the data analysis in the social sciences series. In the Introduction to Social Data you learned the basics of descriptive statistics and R. Data Analysis 1 introduced you to statistical inference. Data Analysis 2 covered linear regression analysis. In Data Analysis 3 we are not going to learn new statistical techniques, but will focus on how to apply the techniques

you already know to the analysis of real-life data sets and how to produce good statistical reports.

This is a skill that you may need in a variety of jobs where data analytic expertise is required, such as marketing analysis, policy analysis in various fields, web analytics, data journalism, academic research, etc.

You already know how to use R to describe data and run simple statistical models. However, real-life data rarely come in the form of a perfectly formatted csv file ready for the analysis. The real life data sets often need to be reshaped, merged, recoded, aggregated and modified in various ways before you can even start your analysis. Unless you know how to do this you will not be able to produce good statistical reports.

This year in this module we will use data from the Understanding Society, a large household panel study conducted in the UK. In the Immigration module we already used the cross-sectional Understanding Society data. In this module we will work with the longitudinal data, which introduces a number of technical challenges.

Throughout the module we will use R for statistical analysis. You are expected to know the basics of data analysis in R.

The only way to learn data analysis is to do data analysis. I will not be able to teach you this, but I can guide your independent learning. This year we will try the "flipped classroom" model of teaching. This means that you will be expected to read and master the required material BEFORE the class and we will use the time in class to answer additional questions and check your solutions rather than introduce new material.

The pre-requisites for this module are POL/SOC1041 and POL/SOC2077.

## 3 Attendance

This module is quite technical. As with other technical skills, missing some initial bits means that you may not be able to catch up. Attendance in this module is crucial. If you do not attend you will not be able to do well in this module. Even skipping a couple of classes will have negative consequences for your understanding of the material. Another negative consequence will be that you will slow the rest of the class down as I will have to explain the same things several times. If you plan not to attend classes please do not take this optional module.

# 4   Assessment

The assessment for this module is a report of 3,500 words (in addition to figures and tables) with the results of statistical analysis you will undertake. This will be 100% of your final mark for this module. You will be given questions for the reports later in the module. In your analysis you will use the Understanding Society data.

The deadline for submitting your reports through eBart is 29 March at 2pm. You will receive your marks and feedback by 5 May.

Late submissions up to two weeks after the deadline will be capped at 40%. Submissions that are late for more than two weeks will not be accepted.

# 5   Syllabus plan

I may change some topics as we proceed.

- Data structures in R

- Manipulating data with dplyr

- Longitudinal data in R. Wide and long formats. Reshaping

- Data visualisation with ggplot2

- Producing statistical reports with R Markdown

- Interactive applications with Shiny

- Loops and other control structures. The apply family of functions

- Writing functions in R

# 6   Reading list

The main text for this module:

- G.Grolemund & H.Wickham. (2016). R for Data Science. Freely available at `http://r4ds.had.co.nz/`.

In addition to this you can the following sources (among many others books on R).

- H.Wickham. (2015). ggplot2. Elegant Graphics for Data Analysis. 2nd ed. Springer.

- W.Chang. (2013). R Graphics Cookbook. O'Reilly.

- P.Spector. (2008). Data Manipulation with R. Springer.

- N.Matloff. (2011). The Art of R Programming. No Starch Press.

- H.Wickham. (2014). Advanced R. Chapman & Hall.